

AMENDMENTS TO THE SPECIFICATION

Page 1, after the title, please insert the following:

--Related Applications

This application is a divisional of U.S. Patent Application Serial No. 09/562,916 filed May 2, 2000 and entitled Construction of Trainable Semantic Vectors and Clustering, Classification, and Searching Using Trainable-Semantic Vectors, which claims priority from U.S. Provisional Patent Application No. 60/177,654, filed January 27, 2000, which is incorporated herein by reference.--

Please amend page 17, line 7 through line 9 as follows:

A weighted average of u and v can also be used to determine the significance of data points, according to the following formula:

$$\text{TSV} = \alpha(v) + (1-\alpha)(u)$$

Please amend page 36, line 2 through line 16 as follows:

Summation of the total number of columns 212 across each row 210 provides the total number of documents that contain the word represented by the row 210. These values are represented at column 216. ~~Summation of all the rows 210 across a column 212 provides the number of documents within the category represented by that column 212. This is shown in Figure 8 using reference numeral 218.~~ Referring to Figure 8 word W₁ appears twenty times in category Cat₂ and eight times in category Cat₅. Word W₁ does not appear in categories Cat₁, Cat₃, and Cat₄. Referring to column 216, word W₁ appears a total of 28 times across all categories. In other words, twenty-eight of the documents classified contain word W₁.

Examination of a an exemplary column 212, such as Cat₁, reveals that word W₂ appears once in category Cat₁, word W₃ appears eight times in category Cat₁, and or W₅ appears twice in category Cat₁. Word W₄ does not appear at all in category Cat₁. As previously stated word W₁ does not appear in category 1. Referring to row 218, the entry corresponding to category Cat₁ indicates that there are eleven documents classified in category Cat₁.

Please amend page 36, line 17 through page 37, line 4 as follows:

With continued reference to Figure 8, Figure 9 illustrates a table 230 that stores the values that indicate the relative strength of each word with respect to the categories. Specifically, the percentage of data points occurring in each category (i.e., u) is presented in the form of a vector for each word. The value for each entry in the u vector is calculated according to the following formula:

$$u = \text{Prob}(\text{entry} | \text{category}) = (\text{entry}_n, \text{category}_m) / \text{category}_{m_total}$$

Table 230 also presents the probability distribution of a data point's occurrence across all categories (i.e., v) in the form of a vector for each word. The value for each entry in the v vector is calculated according to the following formula:

$$v = \text{Prob}(\text{category} | \text{entry}) = (\text{entry} \underline{\text{entry}}_n, \text{category}_m) / \text{entry}_{n_total}$$